On the position of interrogative phrases and the order of complementizer and clause

Matthew S. Dryer

Earlier versions of generative grammar, dating back to Bresnan (1970), proposed that wh-movement moves interrogative phrases into the position of complementizers. While the dominant view in generative grammar since Chomsky (1986) has been that wh-movement is movement into Spec of CP, the purpose of this paper is to examine typological evidence bearing on the earlier view, of movement into the position of complementizers. It investigates crosslinguistic patterns in the position of complementizers and the position of wh-phrases to determine whether there is any correlation between the two. While this does not appear to impact the more recent view of movement to Spec of CP, the patterns described here are of possible independent interest, both to generative linguists and to typologists.[1] I argue that while the typological evidence initially appears to support the idea of a relationship between these two word order parameters, on more careful consideration, I conclude that there is no evidence of a correlation.

Crosslinguistically, we find some languages which normally place wh-phrases at the beginning of sentences, as in English, while other languages normally leave such phrases in situ (Dryer 2011), as in (1) from Khwarshi (a Daghestanian language spoken in Russia).[2]

(1) Khwarshi (Khalilova 2009: 461)

obut-t'-i	uža-l	hibo	b-ez-i?
father-OBL-ERG	boy.OBL-LAT	what	III-buy-PAST.WITNESSED

'What did the father buy his son?'

There is a third type of language that usually places wh-phrases at the beginning of sentences, though it is apparently optional. For example, Curnow (1997) reports that interrogative phrases are normally initial in Awa Pit, a Barbacoan language of Ecuador and Colombia, as in (2a), but occasionally are non-initial, as in (2b).

---

[1] In am indebted to Guglielmo Cinque for helping clarify for me the claims of generative grammar relevant to this paper.

[2] In some verb-final languages, especially in Asia, but apparently not common outside of Asia, it is common for wh-phrases to appear immediately before the verb. In this paper, I will treat such languages as languages with in-situ wh-phrases. In effect, what I refer to as languages with in-situ wh-phrases are simply languages which do not have a rule that normally places wh-phrases at the beginning of sentences.

(2) Awa Pit (Curnow 1997: 315, 316)

a. mɨn=pa kwizha=ta=ma comida kwin-ta-w?
who=POSS dog=ACC=Q food give-PAST-LOCUT:SUBJ
'Whose dog did you give food to?'

b. Libardo=na mɨn-a=ma ta-zi?
Libardo=TOPIC who-ACC=Q give-NONLOCUT
'Who did Libardo pay?'

In the remainder of this paper, I will collapse languages in which wh-phrases are obligatorily or almost always initial with languages like Awa Pit, in which wh-phrases are usually initial, since I assume that for both types of languages, generative linguists would posit a rule of overt wh-movement.[3]

We also find crosslinguistic variation in the position of complementizers. While there are languages which place complementizers at the beginning of clauses, as in English, there are other languages which place complementizers at the end of clauses, as in (3) from Canela (a Ge-Kaingang language spoken in Brazil).

(3) Canela (Popjes and Popjes 1986: 165)

cu-te i-mã amji jarẽ [cu-mã a-kĩn na]
3-PAST 1-TEMPRY self told 3-TEMP 2-like COMP
'He told me that he likes you.'

It should be noted that many languages do not employ complementizers (by which I mean separate words marking complement clauses), either using finite clauses but without a complementizer, as in Begak (an Austronesian language of Sabah), illustrated in (4), or some sort of nominalization, as in Hup (a Nadahup language of Brazil), illustrated in (5), where the subordinate verb is marked with the nominalizer *-n'ɨ̆h*.

(4) Begak (Goudswaard 2005: 338)

K-ingog ku [ikow pədtos].
ACTOR.NONVOLUT-hear 1SG.GEN 2SG.NOM ill
'I heard that you were ill.'

(5) Hup (Epps 2008: 850)

ʔãh hipãh-níh [naw ʔam ʔɨd-n'ɨ̆h]=ɨ́.
1SG know-NEG good 2SG speak-NMZ=DECL
'I didn't know you spoke (Portuguese) so well!'

Under the view that wh-movement is normally movement into complementizer position, we might expect to find a crosslinguistic relationship between the position of

[3] This is in contrast to Dryer (2011), in which languages like Awa Pit are grouped with languages without obligatory initial wh-phrases. In this paper, it makes more sense to group them with languages with in-situ wh-phrases since they might be analysed by generative linguists as having optional wh-movement.

wh-phrases and the position of complementizers. Consider the four logically possible types of languages in (6) defined by the two values given in the preceding paragraph for these two typological parameters.

(6) a. Languages with clause-initial complementizers and in situ wh-phrases
b. Languages with clause-initial complementizers and initial wh-phrases
c. Languages with clause-final complementizers and in situ wh-phrases
d. Languages with clause-final complementizers and initial wh-phrases

If wh-movement is normally movement into complementizer position, then this makes a prediction about the relative frequency of the four language types in (6). Namely, it predicts that we should find few if any languages of type (6d): if complementizer position in a language is clause-final, then there is no complementizer position at the beginning of sentences for wh-phrases to move into. There clearly are ways to get around this if there are languages of this sort, but we would still expect to find somewhat fewer languages of this sort. More precisely, we might expect the ratio of languages of type (6c) to languages of type (6d) to be higher than the ratio of languages of type (6a) to languages of type (6b). In other words, we would expect to find languages with initial wh-phrases to be proportionally more common among languages with initial complementizers than among languages with final complementizers.

Evidence is presented in this paper, based on my current typological database, that this prediction is borne out at best weakly.[4] The relevant numbers of languages are given in Table 1.[5]

CompS & In-Situ-Wh	108
CompS & Initial-Wh	61
SComp & In-Situ-Wh	23
SComp & Initial-Wh	4

Table 1
Position of complementizers and position of interrogative phrases

Table 1 shows four instances of SComp&Initial-Wh languages. One of these is Canela; example (3) above illustrates the clause-final complementizer while (7) illustrates an initial wh-phrase (Popjes and Popjes 1986: 153 state explicitly that wh-phrases occur initially).

[4] My current database is an update on the data given in Dryer (2011) and other chapters by myself in Dryer and Haspelmath (2011), but also includes data on some features not represented in the WALS atlas, such as the order of complementizer and clause discussed in this paper.

[5] In some of my other papers (e.g. Dryer 1992), I cite data in terms of genera grouped into continental-sized areas. For reasons of simplicity, I cite data in this paper simply in terms of numbers of languages, though geographical and genealogical factors can skew such numbers. For reasons discussed in detail elsewhere (Dryer 1989, 2009), one cannot apply statistical tests like the Chi-Square test to data in tables like Table 1.

(7) Canela (Popjes and Popjes 1986: 157)

jũm mãri capi cakôc xàte
about what Capi speak NMLZR
'About what did Capi speak?'

A second is Khoekhoe (also known as Nama, a Khoisan language spoken in Namibia); (8a) illustrates a clause-final complementizer, while (8b) and (8c) illustrate the fact that interrogative words are obligatorily initial in Khoekhoe; (8c), with the interrogative word in situ, is reported by Hagman (1977: 142) to be ungrammatical.[6]

(8) a. tsǐ̀i //'ĩipàkxm̀ ke kè mǐ́í-pa !úũkxm̀ ta !xáisà.
and 1DU.MASC DECL REMOTE.PAST tell-APPLIC 1DU.MASC go COMP
'And we told him that we were going.' (Hagman 1977: 138)

b. taré'e=p //'ĩipà kè ≠'ṹi?
what=3SG.MASC 3SG.MASC REMOTE.PAST eat
'What did he eat?' (Hagman 1977: 142)

c. * //'ĩipà taré'e kè ≠'ṹi?
3SG.MASC what REMOTE.PAST eat
'What did he eat?' (Hagman 1977: 142)

The other two SComp Initial-Wh languages are Urarina (Olawsky 2006) and Osage (Quintero 2004).

At first sight, the numbers in Table 1 might seem to support the predictions: languages with clause-final complementizers but initial wh-phrases are clearly by far the least frequent type among the four types, with only four cases. However, this is due, at least in part, to the relative frequency among each of the two pairs of typological parameters underlying the typology. First, CompS languages are far more common than SComp languages, outnumbering them by 169 to 27 in Table 1. Second, languages with in situ wh-phrases are about twice as common as languages which normally place wh-phrases in initial position, outnumbering them by 131 to 65. Since SComp&Initial-Wh languages have the less common value for both of these parameters,

---

[6] It would be of interest to investigate the relationship between the position of wh-phrases in embedded questions and the position of complementizers. Unfortunately I have very little data on this question. However, the following example from Khoekhoe illustrates an embedded question with *mã̀apá* 'where' at the beginning of the clause and the complementizer *!xáisà* at the end of the clause.

(i) siíkxm̀ ke //'ĩiku tsǐ̀i-à mǐ́í-pa tama kè
1DU.EXCL.MASC DECL 3PL.MASC either-SUBORD tell-APPLIC NEG REMOTE.PAST

hã̀a 'ií mã̀apá=kxm̀ ta !úū !xáisà
PERF PAST where=1DU.MASC IMPERF go COMP

'We didn't tell them either where we were going.' (Hagman 1977: 142)

we would expect them to be the least common of the four types, even if there is no relationship between the position of complementizers and whether a language has initial wh-phrases.

On the other hand, while we might expect SComp&Initial-Wh languages to be the least common of the four types even if there is relationship between the two parameters, we might still expect them to be more common than they are. If the ratio of In-Situ-Wh languages to Initial-Wh languages were the same among SComp languages as it is among CompS languages, then since the ratio among CompS languages is 116 to 53 or about 2.19 to 1, we might expect the distribution among the 27 SComp languages to be more like 19 In-Situ-Wh languages and 8 Initial-Wh languages. But we find only four Initial-Wh languages among the SComp languages, which is only one half of what we might expect if there were no relationship. So perhaps the data in Table 1 does suggest some relationship between these two parameters.

However, the situation is more complicated than this. The order of complementizer and clause and the position of wh-phrases both correlate with the order of object and verb. The data in Table 2 provides clear evidence of a relationship between the order of object and verb and the order of complementizer and clause.

OV&CompS	37
OV&SComp	32
VO&CompS	162
VO&SComp	1

Table 2
Order of object and verb and position of complementizers

Table 2 shows that among OV languages, the two orders of complementizer and clause are about equally common (37 CompS and 32 SComp).[7] But among VO languages, my database contains only one instance of an SComp languages: the other 162 VO languages in this sample are CompS.[8]

Table 3 provides evidence of a correlation between the order of object and verb and whether the language employs initial wh-phrases.

OV&In-Situ-Wh	320
OV&Initial-Wh	95
VO&In-Situ-Wh	259
VO&Initial-Wh	167

Table 3
Order of object and verb and position of interrogative phrases

[7] Table 2 shows evidence of another asymmetry among OV and VO languages: complementizers are more common in VO languages than in OV languages, in this data by 163 languages to 69. This asymmetry does not appear to be relevant to this paper.
[8] The sole instance in my database of a VO language with final complementizer is Hkongso, a Tibeto-Burman language of Burma (Wright 2009).

Table 3 shows that among OV languages In-Situ-Wh languages outnumbers Initial-Wh languages by 320 to 95, over 3 to 1. And while In-Situ-Wh outnumbers Initial-Wh among VO languages as well, by 259 to 167, the ratio is much less than with OV languages. The same point can be made perhaps even more clearly by comparing In-Situ-Wh languages with Initial-Wh languages. Among In-Situ-Wh languages, OV is slightly more common than VO (by 318 to 259). But among Initial-Wh languages, VO is more common by 166 to 95 (or almost 2 to 1). In short, apart from the skewing that results from the fact that In-Situ-Wh is more common than Initial-Wh, we find a correlation between VO and Initial-Wh and between OV and In-Situ-Wh[9].

The fact that both the order of complementizer and clause and the position of wh-phrases correlate with the order of object and verb leads to a reinterpretation of the data in Table 1. The fact that SComp and In-Situ-Wh are each associated with OV means that we expect more SComp&InSituWh languages simply due to the fact that both values are associated with OV. And the fact that CompS and Initial-Wh are each associated with VO means that we expect more CompS&Initial-Wh languages simply due to the fact that both values are associated with VO. Hence the correlations with the order of object and verb favour CompS&Initial-Wh and SComp&InSituWh and disfavour CompS&InSituWh and SComp&Initial-Wh. Table 4 repeats the data from Table 1, but indicates the two types that are favoured by the correlations with the order of object and verb.

CompS & In-Situ-Wh	108	
CompS & Initial-Wh	61	favoured, since both types are associated with VO
SComp & In-Situ-Wh	23	favoured, since both types are associated with OV
SComp & Initial-Wh	4	

Table 4
Position of complementizers and position of interrogative phrases

Earlier in this paper, I suggested that the fact that the ratio of the third to the fourth line in Table 1/4 is greater than the ratio of the first to the second line seems to provide weak support for the claim that there is a crosslinguistic relationship between the order of complementizer and clause and the position of wh-phrases. However, we now have an alternative explanation for this fact: the correlations with the order of object and verb favour two types and this leads us to expect the ratio of the third to the fourth line in Table 4 to be greater than the ratio of the first to the second line. Hence, there is no reason to interpret the data in Tables 1 and 4 as providing weak support for the idea that there is crosslinguistic relationship between wh-movement and complementizer position.

The same point can be made in another way. Table 5 elaborates on the data in Tables 1 and 4, by restricting attention to OV languages. (Looking at VO languages would be unhelpful since almost all the VO languages in my sample are CompS.)

---

[9] The relationship between the position of wh-phrases and order of object and verb is implied by Universal 12 of Greenberg (1963): "If a language has dominant order VSO in declarative sentences, it always puts interrogative words or phrases first in interrogative word questions; if it has dominant order SOV in declarative sentences, there is never such an invariant rule."

OV & CompS & In-Situ-Wh	30
OV & CompS & Initial-Wh	2
OV & SComp & In-Situ-Wh	22
OV & SComp & Initial-Wh	4

Table 5
Position of complementizers and position of interrogative phrases in OV languages

Table 5 shows a different pattern from Table 1. It is now the case that the ratio of the third to the fourth line (22 to 4) is less than the ratio of the first to the second line (30 to 2). In other words (although the difference in numbers is small), OV languages with Initial-Wh are *more* likely than languages with In-Situ-Wh to be SComp, exactly the opposite of what we might expect if there is a crosslinguistic relationship between wh-movement and complementizer position.

Now one might argue that the infrequency of OV&CompS&Initial-Wh languages is simply due to the fact that CompS and Initial-Wh are both associated with VO word order, so we would not expect to find many such languages among OV languages. However, the data in Table 2 shows that the two orders of complementizer and clause are about equally common in OV languages, CompS being slightly more common. Thus, if there were a relationship between the position of complementizers and whether wh-phrases are initial, we would expect to find more Initial-Wh languages among OV&CompS languages than among OV&SComp languages. But we don't.

My conclusion is that the crosslinguistic evidence does not support the idea that there is a crosslinguistic relationship between wh-movement and complementizer position. This is apparently unproblematic under the view that wh-movement is movement to Spec of CP, but unexpected under the older view that wh-movement is movement to Comp.

References

Bresnan, J. W. 1970. On complementizers: toward a syntactic theory of complement types. *Foundations of Language* 6. 297-321.

Chomsky, Noam (1986). *Barriers*. Linguistic Inquiry Monograph 13. MIT Press.

Curnow, Timothy Jowan (1997). *A Grammar of Awa Pit (Cuaiquer): An Indigenous Language of South-Western Colombia*. Australian National University doctoral thesis.

Dryer, Matthew S. 1989. Large linguistic areas and language sampling. *Studies in Language* 13: 257-292.

Dryer, Matthew S. 1992. The Greenbergian word order correlations. *Language* 68: 81-138.

Dryer, Matthew S. 2009. Problems testing typological correlations with the online WALS. *Linguistic Typology* 13: 121 - 135.

Dryer, Matthew S. 2011. Position of interrogative phrases in content questions. In *The World Atlas of Language Structures online*, Matthew S. Dryer and Martin Haspelmath (eds.). Munich: Max Planck Digital Library, chapter 93. Available online at http://wals.info/chapter/93.

Dryer, Matthew S. and Martin Haspelmath, eds. 2011. *The World Atlas of Language Structures online*. Munich: Max Planck Digital Library. Available online at http://wals.info/.

Epps, Patience. 2008. *A Grammar of Hup*. Berlin: De Gruyter.

Goudswaard, Nelleke. 2005. The *Begak* (Ida'an) Language of Sabah. LOT, Netherlands Graduate School of Linguistics doctoral thesis.

Greenberg, Joseph. 1963. Some universals of grammar with particular reference to the order of meaningful elements. In *Universals of Language*, Joseph H. Greenberg (ed.). 73–113. Second Edition. Cambridge, MA: MIT Press.

Hagman, Roy Stephen. 1977. *Nama Hottentot Grammar*. Indiana University Press.

Khalilova, Zaira. 2009. *A Grammar of Khwarshi.* LOT, Netherlands Graduate School of Linguistics doctoral thesis.

Olawsky, Knut J. 2006. *A Grammar of Urarina*. Berlin: De Gruyter.

Popjes, Jack and Jo Popjes. 1986. Canela-Krahô. In *Handbook of Amazonian Languages 1*. Derbyshire, Desmond C. and Geoffrey K. Pullum (eds.), pp. 128-199.

Quintero, Carolyn. 2004. *The Osage Language*. Lincoln, University of Nebraska Press.

Wright, Jonathan Michael. 2009. *Hkongso Grammar Sketch.* Graduate Institute of Applied Linguistics MA Thesis.

University at Buffalo